Classification of clear cell renal cell carcinoma based on *PKM* alternative splicing

Xiangyu Li[1], Beste Turanli[2], Kajetan Juszczak[1], Woonghee Kim[1], Muhammad Arif[1], Yusuke Sato[3,4], Seishi Ogawa[3,5], Hasan Turkez[6], Jens Nielsen[7], Jan Boren[8], Mathias Uhlen[1], Cheng Zhang[1,9,*], Adil Mardinoglu[1,10,*]

[1]Science for Life Laboratory, KTH - Royal Institute of Technology, Stockholm, Sweden

[2]Department of Bioengineering, Istanbul Medeniyet University, Istanbul, Turkey

[3]Department of Pathology and Tumor Biology, Institute for the Advanced Study of Human Biology (WPI-ASHBi), Kyoto University, Kyoto, Japan

[4]Department of Urology, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan

[5]Department of Medicine, Centre for Hematology and Regenerative Medicine, Karolinska Institute, Stockholm, Sweden

[6]Department of Molecular Biology and Genetics, Erzurum Technical University, Erzurum 25240, Turkey.

[7]Department of Biology and Biological Engineering, Chalmers University of Technology, Gothenburg, Sweden

[8]Department of Molecular and Clinical Medicine, University of Gothenburg and Sahlgrenska University Hospital Gothenburg, Sweden

[9]School of Pharmaceutical Sciences, Zhengzhou University, Zhengzhou, PR China

[10]Centre for Host–Microbiome Interactions, Dental Institute, King's College London, London, SE1 9RT, United Kingdom

*Corresponding authors:

Adil Mardinoglu (adilm@scilifelab.se) and **Cheng Zhang** (cheng.zhang@scilifelab.se)

*Lead Contact:

Adil Mardinoglu (adilm@scilifelab.se)

Emails: xiangyu.li@scilifelab.se; bcalimlioglu@gmail.com; juszczak.kajetan@gmail.com; woonghee.kim@scilifelab.se; muhammad.arif@scilifelab.se; yusuke_s_310@yahoo.co.jp; sogawa-tky@umin.ac.jp; hasanturkez@gmail.com; nielsenj@chalmers.se; Jan.Boren@wlab.gu.se; mathias.uhlen@scilifelab.se; cheng.zhang@scilifelab.se; adilm@scilifelab.se

Summary

Clear cell renal cell carcinoma (ccRCC) accounts for 70–80% of kidney cancer diagnoses and displays high molecular and histologic heterogeneity. Hence, it is necessary to reveal the underlying molecular mechanisms involved in progression of ccRCC to better stratify the patients and design effective treatment strategies. Here, we analyzed the survival outcome of ccRCC patients as a consequence of the differential expression of four transcript isoforms of the pyruvate kinase muscle type (*PKM*). We first extracted a classification biomarker consisting of eight gene pairs whose within-sample relative expression orderings (REOs) could be used to robustly classify the patients into two groups with distinct molecular characteristics and survival outcomes. Next, we validated our findings in a validation cohort and an independent Japanese ccRCC cohort. We finally performed drug repositioning analysis based on transcriptomic expression profiles of drug-perturbed cancer cell lines and proposed that paracetamol, nizatidine, dimethadione and conessine can be repurposed to treat the patients in one of the subtype of ccRCC whereas chenodeoxycholic acid, fenoterol and hexylcaine can be repurposed to treat the patients in the other subtype.

Keywords: *PKM*, alternative splicing, transcriptomics, biomarker, drug repositioning

Introduction

Clear cell renal cell carcinoma (ccRCC) is the most common subtype among renal cancer (Motzer et al., 2017) and ccRCC shows high inter individual heterogeneity(Ricketts et al., 2018). Thus, it is difficult to predict survival outcomes of patients in clinical practice and design effective therapeutic strategies. Previous studies have already proposed strategies for stratification of ccRCC patients into subgroups based on different genetic and/or transcriptomic characteristics and prognoses of the patients (Brannon et al., 2010; Cancer Genome Atlas Research, 2013; Kosari et al., 2005; Takahashi et al., 2001). However, these studies failed to identify a clinically practical biomarker for classification of the patients at the personalized level or recommend personalized chemotherapy regimens for these patients.

In a recent study, we have found that pyruvate kinase muscle type (*PKM*), an enzyme that is involved in the final step of glycolysis and catalyzes the formation of ATP from ADP as phosphoenolpyruvate undergoes dephosphorylation to pyruvate, plays a very important role in controlling tumor metabolism in ccRCC (Li et al., 2019b). We have also observed that the expression level of four protein-coding transcripts of *PKM*, including ENST00000335181, encoding PKM2 which is the most studied isoform of PKM, as well as ENST00000561609, ENST00000389093 and ENST00000568883 are highly associated with patients' prognoses. Among them, high expression of ENST00000335181 and ENST00000561609 indicate a favorable survival while high expression of ENST00000389093 and ENST00000568883 indicate an unfavorable survival. Moreover, a number of conserved biological functions associated with the progression of ccRCC were oppositely dysregulated by these transcripts. Here, we hypothesized that different molecular subtypes among ccRCC patients may be characterized by the different expression patterns induced by these four prognostic transcripts and biomarkers that may be used in clinical practice can be identified.

Previous studies have proposed transcriptomics-based biomarkers for classification of tumors based on the quantitative measurement of one or multiple signature genes (Fujita et al., 2012; Jones et al., 2005; Klatte et al., 2009; Kosari et al., 2005; Zhao et al., 2006). However, this kind of transcriptional signatures are rarely used in clinical practice due to technological and translational barriers (Winslow et al., 2012). Besides problems in tissue sampling and quality control, an important factor is experimental batch effect which brings high variation of gene expression induced by the different

laboratory conditions and personnel (Guan et al., 2018). To solve these problems, the use of biomarkers based on the within-sample relative expression orderings (REOs) of gene pairs has been proposed (Guo et al., 2018; Qi et al., 2016a; Qi et al., 2016b), which is robust against batch effects, invariant to monotone data normalization (Eddy et al., 2010; Wang et al., 2013) and poor sample preparation (Chen et al., 2017b; Cheng et al., 2017; Liu et al., 2017).

In this study, we used the genes dysregulated by the prognostic transcripts of *PKM* to extract classification biomarker instead of using themselves. Since different transcripts of *PKM* share similar sequence, it may be difficult to design distinct primers to detect their relative abundance when using real-time PCR. Thus, gene pair biomarker is more feasible and practical in clinical diagnosis. We applied REOs-based method to identify classification biomarker for ccRCC by extracting the expression profiles of genes which were consistently negatively dysregulated by the four favorable and unfavorable prognostic transcripts of *PKM*. We developed a REOs-based biomarker using the global gene expression profiling of ccRCC in The Cancer Genome Atlas (TCGA) database and stratified the patients into two subtypes exhibiting different transcriptomic expression patterns and different patient prognosis. We also validated our findings in TCGA database as well as in another independent Japanese cohort. We finally proposed several candidate drugs that can be used in treatment of each subtype based on transcriptomic expression profiles of drug-perturbed cancer cell lines from Connectivity Map 2.0 (CMap2).

Result

Identification of signature gene set associated with four prognostic transcripts of *PKM*

In a recent study, we have found that there are molecular subtypes that could be characterized by the expression of the four prognostic PKM transcripts (Li et al., 2019b). In order to develop a REOs-based biomarker, we identified a signature gene set associated with these four transcripts based on the gene expression profiles of TCGA ccRCC samples. We performed differential expression analysis between the tumor samples from patients with high (top 25%) and low (bottom 25%) expression of each favorable transcript, and identified 2010 consistently significantly (FDR < 1.0e-5) differentially expressed genes (DEGs) for the two favorable transcripts (Figure 1). Similarly, we identified 5469 DEGs consistently significantly (FDR < 1.0e-5) DEGs for the two unfavorable transcripts. We found that the two sets of DEGs has a significant overlap (n=1135; hypergeometric distribution test, p<1.11e-16). We also observed that the concordance score of these overlapped genes is 100%, which means the up-regulated genes associated with high expression of favorable transcripts within these 1135 genes are all down-regulated when the unfavorable transcripts exhibit high expression; and vice versa.

We followed-up survival information from the corresponding patients and found 539 of the 1135 genes (of which 305 and 234 are favorable and unfavorable, respectively) are significantly (univariate Cox model, FDR < 0.01) associated with patients' overall survival (OS). To identify the associated biological functions with these 539 genes, we performed GO term enrichment analysis and observed that these genes are significantly enriched in RNA splicing, RNA catabolic process and nuclear transport pathways (FDR<0.05; Table S1). Therefore, we concluded that these 539 genes may be used as the core signature genes that are associated with the differential alternative splicing of *PKM* among ccRCC patients and may be used for classification of tumor samples.

We calculated the co-expression coefficients between the expression of the 539 signature genes and found two major clusters in which all favorable genes are positively co-expressed while all unfavorable genes are negatively co-expressed in the opposite cluster using the hierarchical clustering (Figure 2A). Based on the expression profiles of these 539 signature genes, we employed consensus clustering to classify

TCGA ccRCC samples into distinct stable groups through repeated subsampling and clustering (Wilkerson and Hayes, 2010). As shown in Figure 2B, we determined an optimum number of two clusters, cluster 1 and 2, based on the lowest proportion of ambiguous clustering (Senbabaoglu et al., 2014). Using survival analysis, we observed the patients whose tumor samples classified in cluster 1 (N = 231) had significantly shorter OS than those classified in cluster 2 (N = 297) with statistical significance (log-rank test, P=6.73e-07; Figure 2C). The result demonstrated that there are two different molecular subtypes in ccRCC with significantly different survival outcomes which are strongly associated with the function of the two favorable and two unfavorable transcripts.

Development of the REOs-based classification biomarker

To identify a biomarker that can be used in the clinical practice, we next focused on development of a REOs-based classification biomarker based on the gene expression profiles of the 539 signature genes. In brief, REOs-based biomarkers employs gene pairs with consistently reversed expression orders between the two molecular groups as indicators, and screens for a minimum combination of these gene pairs that serves as risk indicators for classification. In order to obtain a robust biomarker, we generated 100 training and 100 validation datasets by randomly selecting from TCGA ccRCC cohort and randomly separated the samples into two respective groups with 70% and 30% samples. We identified 171 gene pairs that exhibited consistent reverse gene pairs in all training datasets. We next generated 17100 reverse gene pair combinations with a forward selection procedure and selected a final REOs-based biomarker consisted of eight reverse gene pairs with an optimal mean F-score of 0.9725 in all training datasets. The full screening process are shown in Figure 1 and detailed in Method section.

Within this eight gene pairs, if more than four gene pairs exhibited reversal REOs in a sample, this sample would be classified into the high-risk group; otherwise, this sample would be classified into the low-risk group (Figure 2D). We tested these gene pairs in the 100 validation datasets, and found that these gene pairs also showed a good classification accuracy with a mean F-score 0.9742. We also tested these gene pairs using the complete TCGA cohort, and this biomarker classified 231 samples into high-risk group and 297 samples into low-risk group. Notably, these two groups showed significantly different OS (Figure 2E; log rank test, P=1.69e-07).

Validation of the REOs-based classification biomarker

To validate if these gene pairs can be used as a biomarker for classification of ccRCC samples, we tested these gene pairs in 100 ccRCC samples obtained from an independent Japanese cohort. The biomarker classified 35 samples as high-risk group and 65 samples as low-risk group, and these two groups showed significantly different OS (Figure 2F, P=7.46e-05). We next investigated whether the high and low-risk groups identified in both TCGA and Japanese cohorts exhibited similar biological differences. We extracted the top 20% most significant DEGs (n=2694) between high and low-risk groups in both the TCGA and Japanese cohorts, and observed a significant overlap between them (n =1463; hypergeometric distribution test, p < 1.11e-16) with a concordance score 100%. In addition, we identified 66 and 80 GO terms that are significantly enriched with upregulated genes (FDR<1.0e-05) in the high-risk group of the TCGA and Japanese cohorts, respectively, and found that 55 of them are common in both cohorts (Figure 3). Specifically, the high-risk group was characterized by upregulated genes involved in ATP synthesis, mitochondrial respiratory process, oxidative phosphorylation, ribonucleotide and purine nucleotide metabolic process, RNA catabolic process, protein targeting to ER and membrane pathways. And the low-risk group was characterized by upregulated genes involved in histone modification and covalent chromatin modification pathways. The results suggested the molecular subtypes identified by our analysis also have consistent biological differences. Moreover, these 55 GO terms included all 27 GO terms that we recently reported to be associated with the four prognostic transcripts in pan cancer analysis (Li et al., 2019b). We also identified three GO terms that are significantly enriched with downregulated genes in the high-risk group for both cohorts, and two of them, which are the histone modification and covalent chromatin modification pathways, are common in both groups. These results further indicated that the molecular subtypes stratified by the gene pairs are functionally related to the four prognostic transcripts of *PKM.*

Further, we compared our REOs-based classification with previously reported TCGA (m1 to m4) and ccA/ccB classification schemes (Brannon et al., 2010; Cancer Genome Atlas Research, 2013) (Figure 4). In TCGA cohort, approximately 96% of TCGA m1 tumors were involved in our low-risk group, and m1 group was also reported with the best prognoses in TCGA classification scheme. In addition, 73% of TCGA tumors in

both m2 and m4 subtypes were involved in our high-risk groups, and they were also shown to be with the poorest prognoses in TCGA classification. These results demonstrated that the high and low-risk groups classified by our biomarker are reinforced by the previously observed survival outcomes m1, m2 and m4. Notably, the high and low-risk groups respectively accounted for 42% and 58% of tumors previously reported as unclassified (m3) in the TCGA classification scheme. In Japanese cohort, 71% of ccA and 80% of ccB were observed in the low and high-risk groups, respectively. We found that the favorable survival for ccA cases again reinforced the low-risk group classification based on our gene pairs used as a biomarker (Figure 4).

Drug discovery with reversed expression effect

In addition to classification of the tumors, we also performed drug repositioning analysis to identify drug candidates that can be used in treatment of each subtype. We assumed that if a drug could reverse the dysregulated gene expression pattern from a tumor subtype to normal pattern, it could be potentially useful for treating the specific tumor subtype. We used a method developed in our previous study for drug repurposing (Turanli et al., 2019a; Turanli et al., 2018; Turanli et al., 2019b) and found several drugs that could be used for treatment of the high and low-risk groups. We found that four different drugs including paracetamol, nizatidine, dimethadione and conessine could be used to reverse the gene expression in samples from high-risk group, since over 80% drug-perturbed genes were mapped to the DEGs between these samples and normal samples. Interestingly, it has been reported that paracetamol, an analgesic and antipyretic drug, inhibits the cell proliferation and induces cell apoptosis in pancreatic cancer (Malsy et al., 2017), ovarian cancer and lung cancer cells (Lian et al., 2018). Nizatidine, a histamine H_2 receptor antagonist, was also recommended to be added into the combination therapy for cancer treatment (Barton-Burke, 1996; Ben-Sasson, 2007; Feitelberg et al., 2013). Therefore, the anti-cancer effects of two of the proposed drugs have been validated in previous studies.

Similarly, we found that three different drugs including chenodeoxycholic acid, fenoterol and hexylcaine, could be used to reverse the gene expression in samples of low-risk group towards normal samples. It has been reported that chenodeoxycholic acid, a bile acid, shows anti-proliferative activity in human cancer cells (Faustino et al., 2016). Fenoterol, a β adrenoreceptor agonist, has been shown to inhibit proliferation of glioblastomas and astrocytomas cells (Bernier et al., 2013; Toll et al., 2011).

Hexylcaine, a short-acting local anesthetic, has also been used to treat cancer (Gleich, 2000). In this context, these three drugs may also be potentially used for treatment of the subtype of ccRCC patients.

Moreover, we identified the gene targets for each of the drugs using DrugBank database (Wishart et al., 2006). *HRH2*, encoding the histamine H2 receptor, has been reported as the gene target of nizatidine (Meredith et al., 1985). We observed that it is significantly up-regulated in the samples of high-risk group compared to normal samples. It has been demonstrated that in vitro and in vivo histamine-induced tumor cell proliferation can be blocked by H2 antagonists (Deva and Jameson, 2012; Natori et al., 2005; Tomita et al., 2003). Thus, nizatidine may be used as a promising drug for the patients classified in the high-risk group. On the other hand, *GPBAR1*, encoding an enzyme of the G protein-coupled receptor superfamily, has been reported as a target of chenodeoxycholic acid. It has been shown that *GPBAR1* antagonizes kidney cancer cell proliferation and migration (Su et al., 2017). Based on our analysis, we have observed that *GPBAR1* is significantly downregulated in the low-risk group compared to normal samples. Thus, chenodeoxycholic acid, as an activator of *GPBAR1*, may be used as a promising drug for treating the patients classified in the low-risk group.

Discussion

PKM is one of the important regulators of Warburg effect in different human cancers (Dayton et al., 2016). Our recent study showed that four different transcripts of PKM mediated opposite survival outcomes for ccRCC patients. In this study, we identified the core signature genes which were consistently dysregulated by these four prognostic transcripts of *PKM*. Using these signature genes, we identified eight gene pairs whose within-samples REOs could be used to classify patients into two groups with significantly different OS. REOs-based biomarkers take the advantages of the robustness of the intra sample gene expression pattern, and it is relatively insensitive to both experimental and bioinformatics variations compared to conventional biomarkers based on absolute quantification. Although RNA sequencing data was used for the biomarker classification in this study, much cheaper technics could be used once the biomarker is used in clinical practice. For instance, we could use real-time PCR, which is much cheaper compared to the sequencing approach, to determine the relative abundance of the genes involved in these 8 gene pairs to classify a ccRCC tumor sample since we only need to detect their REOs. This could greatly facilitate the use of REOs-based biomarker in clinical practice.

The genes involved in our classification of ccRCC tumor samples also showed closed relationship with tumor development. For instance, RP9, one of genes involved in the REOs gene pairs, plays an important role in pre-mRNA splicing and could interact with well-known oncogene PIM-1(Maita et al., 2000). TAZ, another gene involved in the REOs gene pairs, encodes tafazzin whose overexpression promotes tumorigenicity in many cancers and its inhibition also induces tumor cell apoptosis (Chen et al., 2017a; Li et al., 2019a; Pathak et al., 2014). Another example is NOLC1 which functions as a chaperone for shuttling between the nucleolus and cytoplasm (Meier and Blobel, 1992). It has been reported that enhancement of NOLC1 promotes cell senescence and represses hepatocellular carcinoma cell proliferation by disturbing the organization of nucleolus (Yuan et al., 2017).

In conclusion, we identified two molecular subtypes of ccRCC patients with high and low-risk of mortality, and developed a REOs-based classification biomarker which could be used to identify which subtype of the patients belong to in a personalized manner. In addition, we also suggested specific treatment strategies for each subtype based on their global gene expression patterns. Therefore, it is worthwhile to further

explore the potential clinical use of the here identified biomarker in assisting clinical diagnosis and treatment of ccRCC patients.

Materials and methods

Data and preprocessing

The TCGA transcript-expression level profiles (TPM and count values) of ccRCC and matched normal kidney samples was downloaded from https://osf.io/gqrz9 (Tatlow and Piccolo, 2016) on November 27, 2018, which was quantified by Kallisto (Bray et al., 2016) based on the GENCODE reference transcriptome (version 24). The clinical information of TCGA samples was downloaded through R package TCGAbiolinks (Colaprico et al., 2016). The whole-exome sequence data of 100 ccRCC samples of patients from Japanese cohort (Sato et al., 2013) was downloaded from European Genome-phenome Archive (accession number: EGAS00001000509). BEDTools (Quinlan and Hall, 2010) was used for converting BAM to FASTQ file. Kallisto was used for estimating the count and TPM values of transcripts based on the same reference transcriptome of TCGA data. The sum value of the multiple transcripts of a gene was used as the expression value of this gene. The genes with average TPM values >1 in ccRCC patients were analyzed.

Differential expression analysis

DESeq2 (Love et al., 2014) was used to identify DEGs between two groups. The raw count values of genes were used as input of DESeq2. The Benjamini-Hochberg (BH) procedure was used to estimate FDR.

Overlapping of two lists of DEGs

If DEG list 1 with L_1 genes and DEG list 2 with L_2 genes have k overlapping genes and s of these genes shows the same directions which means high expression of these genes indicates favorable/unfavorable survival or group 1/2 in both lists, the probability of observing at least s consistent genes by chance can be calculated according to the following cumulative hypergeometric distribution model:

$$\mathrm{P} = 1 - \sum_{i=0}^{s-1} \frac{\binom{L2}{i}\binom{L-L2}{L1-i}}{\binom{L}{L1}}$$

where *L* represents the number of the background genes commonly detected in the datasets from which the DEGs are extracted. The two DEG lists were considered to be significantly overlapping if $P < 0.05$.

The concordance score *s*/*k* is used to evaluate the consistency of DEGs between the two lists. Obviously, the score ranges from 0 to 1, and the higher concordance score suggests the better consistency of two lists of DEGs.

Consensus clustering

Consensus clustering (Wilkerson and Hayes, 2010) was used for tumor classification based on the normalized expression profiles of signature genes by Z-score transformation. To achieve robust clusters, the data was resampled for 1000 times by considering 80% samples and signature genes resampling. The resampled data was transformed into a similarity matrix, termed as consensus matrix. K-means clustering was used to stratify samples based on the consensus matrix. The number of optimum cluster was determined by the lowest proportion of ambiguous clustering.

Survival analysis

The univariate Cox regression model was used to evaluate the correlation of gene expression levels with OS. Survival curves were estimated by the Kaplan-Meier method and compared with the log-rank test.

Functional enrichment analysis

GO enrichment was performed by the enrichGo function in R package ClusterProfiler (Yu et al., 2012), in which the hypergeometric distribution was used to calculate the statistical significance of biological pathways enriched with DEGs of interest.

Development of the REOs-based biomarker

In each sample, the REO of every two signature genes (i and j) is denoted as either $G_i > G_j$ or $G_i < G_j$ exclusively, where G_i and G_j represent the expression values of gene i and j, respectively. For a given gene pair (G_i and G_j), we used Fisher's exact test to evaluate whether the frequency of group 1 samples with a specific REO pattern ($G_i > G_j$ or $G_i < G_j$) was significantly different from that in group 2 samples in each training dataset. The P values are adjusted using BH procedure. The gene pairs detected with 0.05 FDR control and over 70% difference of the frequency of their REOs between two groups were denoted as reversed gene pairs. The overlapped reversed gene pairs

consistently identified from all the training datasets were selected as the candidate signature gene pairs. Totally, we found 171 signature gene pairs. For each signature gene pair, according to their within-sample REO, we classified the samples of each training dataset into high or low-risk groups and then evaluated the sensitivity and specificity of this gene pair. Here, the sensitivity is defined as the ratio of correctly identified high risk samples to all high-risk samples and the specificity is defined as the ratio of correctly identified low-risk samples to all low-risk samples. Then, from these signature gene pairs, we performed a forward selection procedure in each training dataset to search a set of gene pairs that achieved the highest F-score value, a harmonic mean of sensitivity and specificity, which is calculated as follows:

$$F - score = \frac{2(\text{sensitivity} * \text{specificity})}{(\text{sensitivity} + \text{specificity})}$$

Taking each of the 171 gene pairs as a seed, we added another gene pair to the biomarker at a time until the F-score did not increase. The classification rule is that a sample is classified into high or low-risk group if the majority of the REOs of the set of gene pairs within this sample vote for high or low-risk. We got 171 biomarkers based on each training dataset. Totally, we got 17100 candidate biomarkers for all the 100 training datasets. Finally, we selected the biomarker with the lowest $(1\text{-}F_1)^2+(1\text{-}F_2)^2+\cdots+(1\text{-}F_n)^2$ as the final biomarker, in which F_n is the F-score value in the n^{th} training dataset.

Application of CMap2 data to drug discovery

The pre-procession of CMap2 data was described in our previous study (Turanli et al., 2019b). In brief, the gene expression profiles of three cell lines, HL60, MCF2 and PC3, were downloaded from https://portals. broadinstitute.org/cmap/ (CMap Build 02). For each cell line, gene Log2FC was used for comparison between treatment instances and its respective controls. Then, the confidence score was calculated per each drug-gene interaction using the P values from three cell lines. An approximation confidence score to 1 was assumed as the higher confidence level. The drug-gene pairs with CS>0.5 were used in further analysis.

Acknowledgements

The study is funded by The Knut and Alice Wallenberg Foundation.

Declaration of Interests

The authors declare no competing interests.

Reference

Barton-Burke, M. (1996). Cancer chemotherapy: A nursing process approach (Jones & Bartlett Learning).

Ben-Sasson, S.A. (2007). Anti-cancer therapy comprising an H2-blocker, at least one antiinflammatory agent and a cytotoxic agent. United States patent US7838513B2.

Bernier, M., Paul, R.K., Dossou, K.S., Wnorowski, A., Ramamoorthy, A., Paris, A., Moaddel, R., Cloix, J.F., and Wainer, I.W. (2013). Antitumor activity of (R,R')-4-methoxy-1-naphthylfenoterol in a rat C6 glioma xenograft model in the mouse. Pharmacol Res Perspect *1*, e00010.

Brannon, A.R., Reddy, A., Seiler, M., Arreola, A., Moore, D.T., Pruthi, R.S., Wallen, E.M., Nielsen, M.E., Liu, H., Nathanson, K.L*., et al.* (2010). Molecular Stratification of Clear Cell Renal Cell Carcinoma by Consensus Clustering Reveals Distinct Subtypes and Survival Patterns. Genes Cancer *1*, 152-163.

Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol *34*, 525-527.

Cancer Genome Atlas Research, N. (2013). Comprehensive molecular characterization of clear cell renal cell carcinoma. Nature *499*, 43-49.

Chen, M., Zhang, Y., and Zheng, P.S. (2017a). Tafazzin (TAZ) promotes the tumorigenicity of cervical cancer cells and inhibits apoptosis. PLoS One *12*, e0177171.

Chen, R., Guan, Q., Cheng, J., He, J., Liu, H., Cai, H., Hong, G., Zhang, J., Li, N., Ao, L*., et al.* (2017b). Robust transcriptional tumor signatures applicable to both formalin-fixed paraffin-embedded and fresh-frozen samples. Oncotarget *8*, 6652-6662.

Cheng, J., Guo, Y., Gao, Q., Li, H., Yan, H., Li, M., Cai, H., Zheng, W., Li, X., Jiang, W*., et al.* (2017). Circumvent the uncertainty in the applications of transcriptional signatures to tumor tissues sampled from different tumor sites. Oncotarget *8*, 30265-30275.

Colaprico, A., Silva, T.C., Olsen, C., Garofano, L., Cava, C., Garolini, D., Sabedot, T.S., Malta, T.M., Pagnotta, S.M., Castiglioni, I*., et al.* (2016). TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. Nucleic Acids Res *44*, e71.

Dayton, T.L., Jacks, T., and Vander Heiden, M.G. (2016). PKM2, cancer metabolism, and the road ahead. EMBO Rep *17*, 1721-1730.

Deva, S., and Jameson, M. (2012). Histamine type 2 receptor antagonists as adjuvant treatment for resected colorectal cancer. Cochrane Database Syst Rev, CD007814.

Eddy, J.A., Sung, J., Geman, D., and Price, N.D. (2010). Relative expression analysis for molecular cancer diagnosis and prognosis. Technol Cancer Res Treat *9*, 149-159.

Faustino, C., Serafim, C., Rijo, P., and Reis, C.P. (2016). Bile acids and bile acid derivatives: use in drug delivery systems and as therapeutic agents. Expert Opin Drug Deliv *13*, 1133-1148.

Feitelberg, D., Berkman, T., Ben-Sasson, S., and Goldstaub, D. (2013). Combination therapy for the treatment of cancer. United States patent US20150005252A1.

Fujita, T., Iwamura, M., Ishii, D., Tabata, K., Matsumoto, K., Yoshida, K., and Baba, S. (2012). C-reactive protein as a prognostic marker for advanced renal cell carcinoma treated with sunitinib. Int J Urol *19*, 908-913.

Gleich, G.J. (2000). Topical anesthetics useful for treating cancer. United States patent US6391888B1.

Guan, Q., Yan, H., Chen, Y., Zheng, B., Cai, H., He, J., Song, K., Guo, Y., Ao, L., Liu, H*., et al.* (2018). Quantitative or qualitative transcriptional diagnostic signatures? A case study for colorectal cancer. BMC Genomics *19*, 99.

Guo, Y., Jiang, W., Ao, L., Song, K., Chen, H., Guan, Q., Gao, Q., Cheng, J., Liu, H., Wang, X*., et al.* (2018). A qualitative signature for predicting pathological response to neoadjuvant chemoradiation in locally advanced rectal cancers. Radiother Oncol *129*, 149-153.

Jones, J., Otu, H., Spentzos, D., Kolia, S., Inan, M., Beecken, W.D., Fellbaum, C., Gu, X., Joseph, M., Pantuck, A.J*., et al.* (2005). Gene signatures of progression and metastasis in renal cell cancer. Clin Cancer Res *11*, 5730-5739.

Klatte, T., Seligson, D.B., LaRochelle, J., Shuch, B., Said, J.W., Riggs, S.B., Zomorodian, N., Kabbinavar, F.F., Pantuck, A.J., and Belldegrun, A.S. (2009). Molecular signatures of localized clear cell renal cell carcinoma to predict disease-free survival after nephrectomy. Cancer Epidemiol Biomarkers Prev *18*, 894-900.

Kosari, F., Parker, A.S., Kube, D.M., Lohse, C.M., Leibovich, B.C., Blute, M.L., Cheville, J.C., and Vasmatzis, G. (2005). Clear cell renal cell carcinoma: gene expression analyses identify a potential signature for tumor aggressiveness. Clin Cancer Res *11*, 5128-5139.

Li, X., Wu, M., An, D., Yuan, H., Li, Z., Song, Y., and Liu, Z. (2019a). Suppression of Tafazzin promotes thyroid cancer apoptosis via activating the JNK signaling pathway and enhancing INF2-mediated mitochondrial fission. J Cell Physiol.

Li, X., Zhang, C., Kim, W., Arif, M., Gao, C., Hober, A., Kotol, D., Strandberg, L., Forsström1, B., Sivertsson, Å*., et al.* (2019b). Discovery of functional alternatively spliced PKM transcripts in human cancers. bioRXIV, https://www.biorxiv.org/content/10.1101/613364v613361.

Lian, X., Huang, Y., Zhu, Y., Fang, Y., Zhao, R., Joseph, E., Li, J., Pellois, J.P., and Zhou, H.C. (2018). Enzyme-MOF Nanoreactor Activates Nontoxic Paracetamol for Cancer Therapy. Angew Chem Int Ed Engl *57*, 5725-5730.

Liu, H., Li, Y., He, J., Guan, Q., Chen, R., Yan, H., Zheng, W., Song, K., Cai, H., Guo, Y*., et al.* (2017). Robust transcriptional signatures for low-input RNA samples based on relative expression orderings. BMC Genomics *18*, 913.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol *15*, 550.

Maita, H., Harada, Y., Nagakubo, D., Kitaura, H., Ikeda, M., Tamai, K., Takahashi, K., Ariga, H., and Iguchi-Ariga, S.M. (2000). PAP-1, a novel target protein of phosphorylation by pim-1 kinase. Eur J Biochem *267*, 5168-5178.

Malsy, M., Graf, B., and Bundscherer, A. (2017). Effects of metamizole, MAA, and paracetamol on proliferation, apoptosis, and necrosis in the pancreatic cancer cell lines PaTu 8988 t and Panc-1. BMC Pharmacol Toxicol *18*, 77.

Meier, U.T., and Blobel, G. (1992). Nopp140 shuttles on tracks between nucleolus and cytoplasm. Cell *70*, 127-138.

Meredith, C.G., Speeg, K.V., Jr., and Schenker, S. (1985). Nizatidine, a new histamine H2-receptor antagonist, and hepatic oxidative drug metabolism in the rat: a comparison with structurally related compounds. Toxicol Appl Pharmacol *77*, 315-324.

Motzer, R.J., Jonasch, E., Agarwal, N., Bhayani, S., Bro, W.P., Chang, S.S., Choueiri, T.K., Costello, B.A., Derweesh, I.H., Fishman, M*., et al.* (2017). Kidney Cancer, Version 2.2017, NCCN Clinical Practice Guidelines in Oncology. J Natl Compr Canc Netw *15*, 804-834.

Natori, T., Sata, M., Nagai, R., and Makuuchi, M. (2005). Cimetidine inhibits angiogenesis and suppresses tumor growth. Biomed Pharmacother *59*, 56-60.

Pathak, S., Meng, W.J., Zhang, H., Gnosa, S., Nandy, S.K., Adell, G., Holmlund, B., and Sun, X.F. (2014). Tafazzin protein expression is associated with tumorigenesis and radiation response in rectal cancer: a study of Swedish clinical trial on preoperative radiotherapy. PLoS One *9*, e98317.

Qi, L., Chen, L., Li, Y., Qin, Y., Pan, R., Zhao, W., Gu, Y., Wang, H., Wang, R., Chen, X*., et al.* (2016a). Critical limitations of prognostic signatures based on risk scores summarized from gene expression levels: a case study for resected stage I non-small-cell lung cancer. Brief Bioinform *17*, 233-242.

Qi, L., Li, Y., Qin, Y., Shi, G., Li, T., Wang, J., Chen, L., Gu, Y., Zhao, W., and Guo, Z. (2016b). An individualised signature for predicting response with concordant survival benefit for lung adenocarcinoma patients receiving platinum-based chemotherapy. Br J Cancer *115*, 1513-1519.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841-842.

Ricketts, C.J., De Cubas, A.A., Fan, H., Smith, C.C., Lang, M., Reznik, E., Bowlby, R., Gibb, E.A., Akbani, R., Beroukhim, R.*, et al.* (2018). The Cancer Genome Atlas Comprehensive Molecular Characterization of Renal Cell Carcinoma. Cell Rep *23*, 3698.

Sato, Y., Yoshizato, T., Shiraishi, Y., Maekawa, S., Okuno, Y., Kamura, T., Shimamura, T., Sato-Otsubo, A., Nagae, G., Suzuki, H.*, et al.* (2013). Integrated molecular analysis of clear-cell renal cell carcinoma. Nat Genet *45*, 860-867.

Senbabaoglu, Y., Michailidis, G., and Li, J.Z. (2014). Critical limitations of consensus clustering in class discovery. Sci Rep *4*, 6207.

Su, J., Zhang, Q., Qi, H., Wu, L., Li, Y., Yu, D., Huang, W., Chen, W.D., and Wang, Y.D. (2017). The G-protein-coupled bile acid receptor Gpbar1 (TGR5) protects against renal inflammation and renal cancer cell proliferation and migration through antagonizing NF-kappaB and STAT3 signaling pathways. Oncotarget *8*, 54378-54387.

Takahashi, M., Rhodes, D.R., Furge, K.A., Kanayama, H., Kagawa, S., Haab, B.B., and Teh, B.T. (2001). Gene expression profiling of clear cell renal cell carcinoma: gene identification and prognostic classification. Proc Natl Acad Sci U S A *98*, 9754-9759.

Tatlow, P.J., and Piccolo, S.R. (2016). A cloud-based workflow to quantify transcript-expression levels in public cancer compendia. Sci Rep *6*, 39259.

Toll, L., Jimenez, L., Waleh, N., Jozwiak, K., Woo, A.Y., Xiao, R.P., Bernier, M., and Wainer, I.W. (2011). {Beta}2-adrenergic receptor agonists inhibit the proliferation of 1321N1 astrocytoma cells. J Pharmacol Exp Ther *336*, 524-532.

Tomita, K., Izumi, K., and Okabe, S. (2003). Roxatidine- and cimetidine-induced angiogenesis inhibition suppresses growth of colon cancer implants in syngeneic mice. J Pharmacol Sci *93*, 321-330.

Turanli, B., Altay, O., Boren, J., Turkez, H., Nielsen, J., Uhlen, M., Arga, K.Y., and Mardinoglu, A. (2019a). Systems biology based drug repositioning for development of cancer therapy. Semin Cancer Biol.

Turanli, B., Karagoz, K., Gulfidan, G., Sinha, R., Mardinoglu, A., and Arga, K.Y. (2018). A Network-Based Cancer Drug Discovery: From Integrated Multi-Omics Approaches to Precision Medicine. Curr Pharm Des *24*, 3778-3790.

Turanli, B., Zhang, C., Kim, W., Benfeitas, R., Uhlen, M., Arga, K.Y., and Mardinoglu, A. (2019b). Discovery of therapeutic agents for prostate cancer using genome-scale metabolic modeling and drug repositioning. EBioMedicine *42*, 386-396.

Wang, H., Zhang, H., Dai, Z., Chen, M.S., and Yuan, Z. (2013). TSG: a new algorithm for binary and multi-class cancer classification and informative genes selection. BMC Med Genomics *6 Suppl 1*, S3.

Wilkerson, M.D., and Hayes, D.N. (2010). ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics *26*, 1572-1573.

Winslow, R.L., Trayanova, N., Geman, D., and Miller, M.I. (2012). Computational medicine: translating models to clinical care. Sci Transl Med *4*, 158rv111.

Wishart, D.S., Knox, C., Guo, A.C., Shrivastava, S., Hassanali, M., Stothard, P., Chang, Z., and Woolsey, J. (2006). DrugBank: a comprehensive resource for in silico drug discovery and exploration. Nucleic Acids Res *34*, D668-672.

Yu, G., Wang, L.G., Han, Y., and He, Q.Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS *16*, 284-287.

Yuan, F., Zhang, Y., Ma, L., Cheng, Q., Li, G., and Tong, T. (2017). Enhanced NOLC1 promotes cell senescence and represses hepatocellular carcinoma cell proliferation by disturbing the organization of nucleolus. Aging Cell *16*, 726-737.

Zhao, H., Ljungberg, B., Grankvist, K., Rasmuson, T., Tibshirani, R., and Brooks, J.D. (2006). Gene expression profiling predicts survival in conventional renal cell carcinoma. PLoS Med *3*, e13.

Figure legends

Figure 1. The flowchart for developing and validating the ccRCC classification biomarker. In brief, we extract the 1135 overlapped DEGs associated with favorable and unfavorable transcripts, and select 539 prognostic signature genes from them. Next, we screen gene pair biomarker using randomly generated training dataset. Lastly, we validate the performance of the biomarker in all randomly generated validation dataset and an independent Japanese ccRCC dataset.

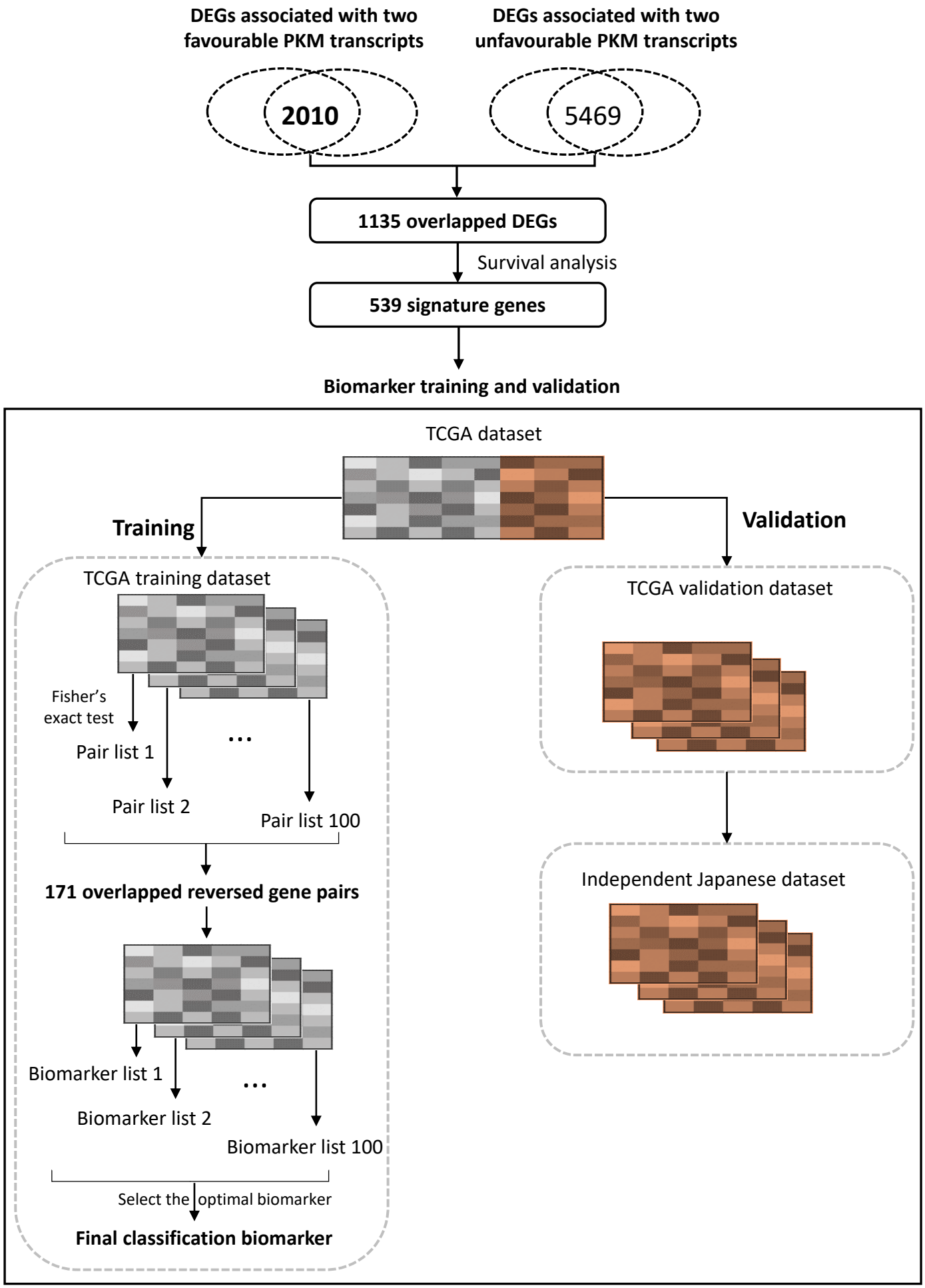
Figure 2. Molecular classification and prognostic prediction of patients by classification biomarker. (A) Hierarchical clustering of 539 signature genes based on the correlation between genes. The spearman correlation coefficients between genes were used for clustering. (B) Consensus clustering for TCGA ccRCC patients based on the expression values of the 539 signature genes. (C) Kaplan-Meier plot of OS of two clusters identified by consensus clustering in TCGA ccRCC cohort. (D) The composition of classification biomarker and voting rule. (E) Kaplan-Meier plot of OS of high- and low-risk identified by classification biomarker in TCGA ccRCC cohort. (F) Kaplan-Meier plot of OS of high- and low-risk identified by classification biomarker in Japanese ccRCC cohort.

Figure 3. The dysregulated biological functions in high- and low-risk ccRCC groups and. Heat map of the p values (on the negative log 10 scale) for the enriched GO terms in TCGA and Japanese KIRC cohort. Red color denotes the GO terms enriched with up-regulated genes. Blue color denotes the GO terms enriched with down-regulated genes. * FDR<1.0e-05.

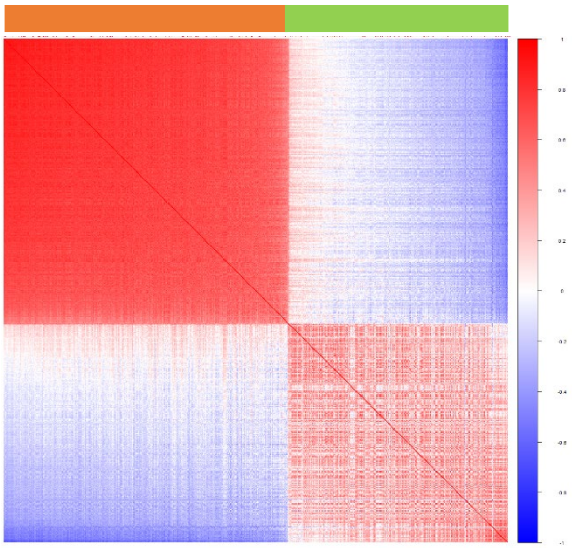
Figure 4. Pie charts showing the intersection of the different classification schemes for ccRCC. ‘m1’, ‘m2’, ‘m3’ and ‘m4’ indicate the molecular subtypes proposed by TCGA, and ‘ccA’ and ‘ccB’ are molecular subtypes reported by another previous study.

Supplementary Table legend

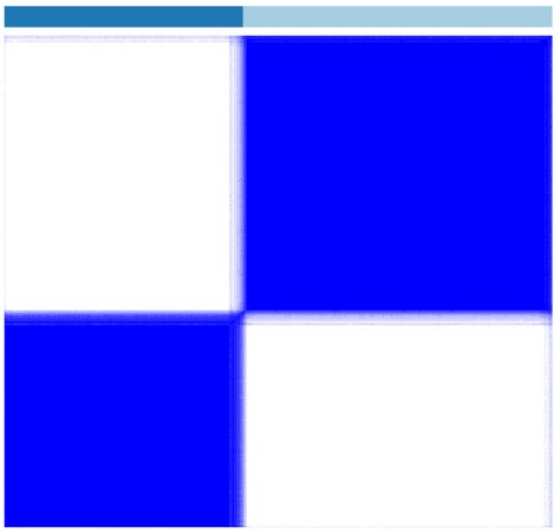
Table S1. The enriched GO terms for the 539 signature genes


DEGs associated with two favourable PKM transcripts
DEGs associated with two unfavourable PKM transcripts
2010
5469
1135 overlapped DEGs
Survival analysis
539 signature genes
Biomarker training and validation
TCGA dataset
Training
Validation
TCGA training dataset
Fisher's exact test
Pair list 1
Pair list 2
Pair list 100
171 overlapped reversed gene pairs
Biomarker list 1
Biomarker list 2
Biomarker list 100
Select the optimal biomarker
Final classification biomarker
TCGA validation dataset
Independent Japanese dataset

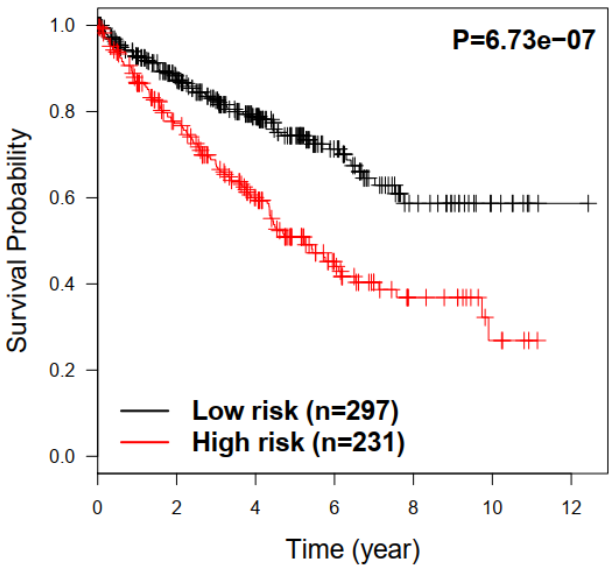
A



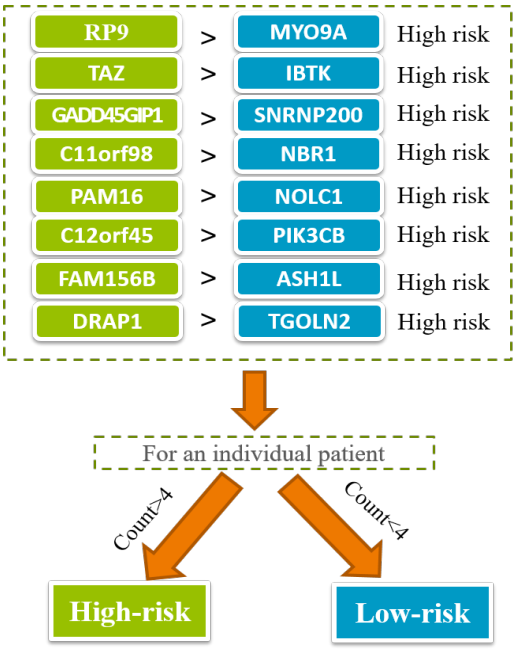
B



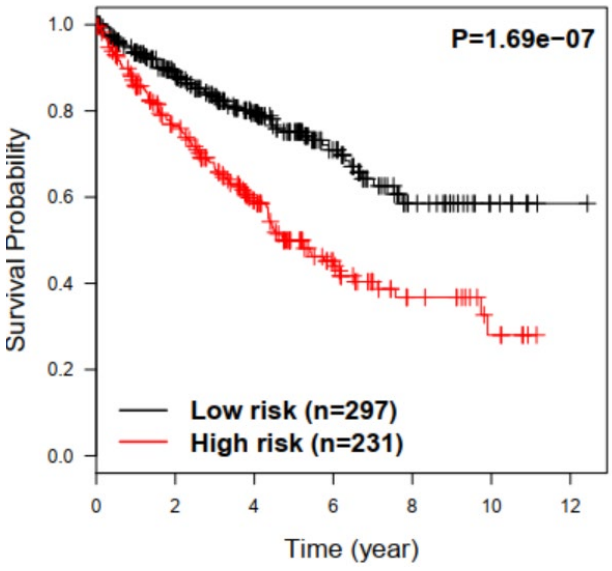
C


P=6.73e−07
Survival Probability
Time (year)
Low risk (n=297)
High risk (n=231)

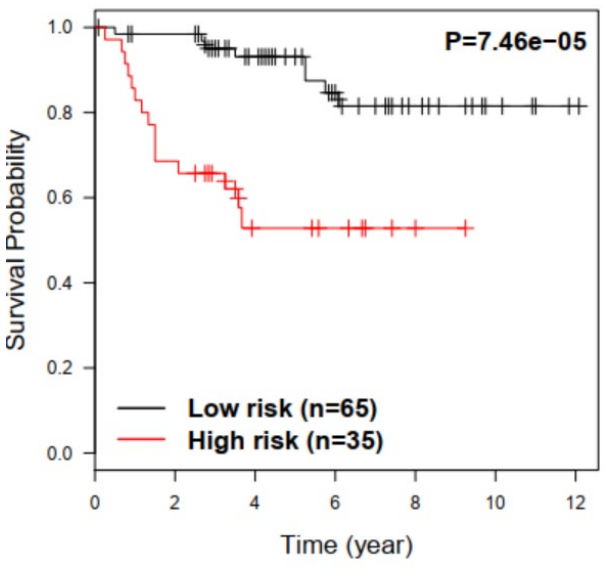
D

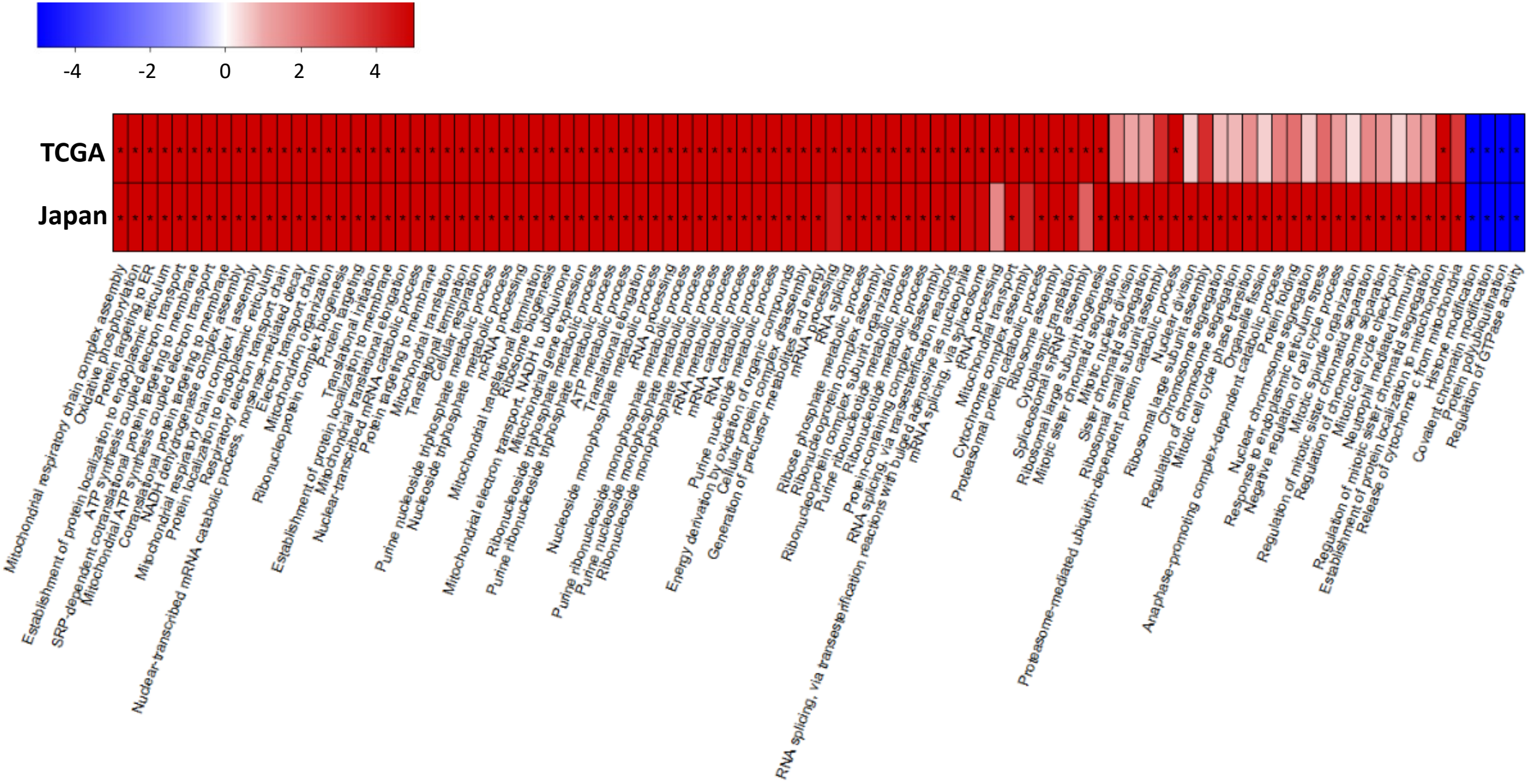

RP9 > MYO9A High risk
TAZ > IBTK High risk
GADD45GIP1 > SNRNP200 High risk
C11orf98 > NBR1 High risk
PAM16 > NOLC1 High risk
C12orf45 > PIK3CB High risk
FAM156B > ASH1L High risk
DRAP1 > TGOLN2 High risk
For an individual patient
Count≥4
Count<4
High-risk
Low-risk

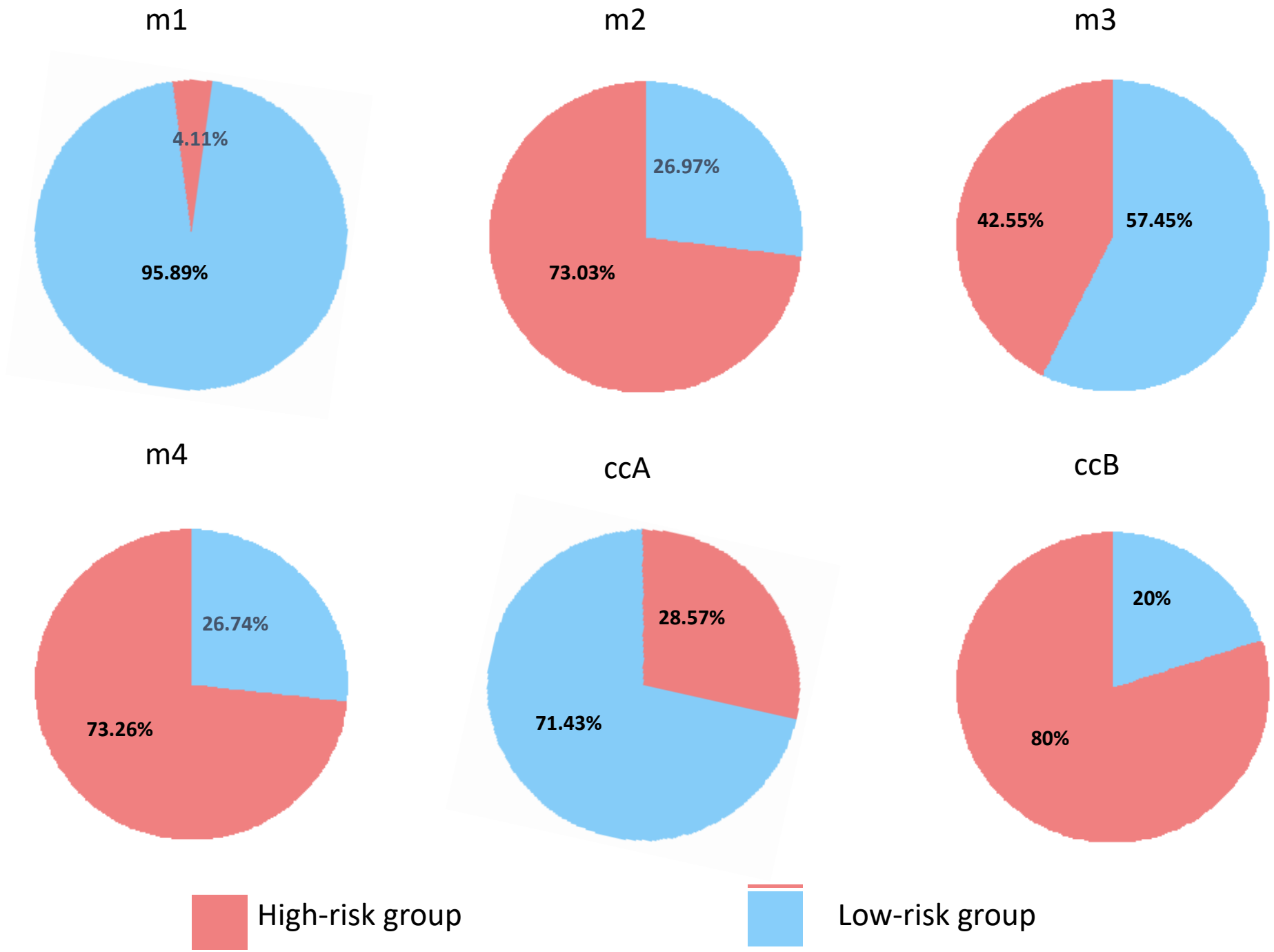
E


P=1.69e−07
Survival Probability
Time (year)
Low risk (n=297)
High risk (n=231)

F


P=7.46e−05
Survival Probability
Time (year)
Low risk (n=65)
High risk (n=35)


-4
-2
0
2
4
TCGA
Japan
Mitochondrial respiratory chain complex assembly
Establishment of protein localization to endoplasmic reticulum
Oxidative phosphorylation
Protein targeting to ER
SRP-dependent cotranslational protein targeting to membrane
ATP synthesis coupled electron transport
Mitochondrial ATP synthesis coupled electron transport
Cotranslational protein targeting to membrane
NADH dehydrogenase complex assembly
Mitochondrial respiratory chain complex I assembly
Protein localization to endoplasmic reticulum
Respiratory electron transport
Nuclear-transcribed mRNA catabolic process, nonsense-mediated decay
Electron transport chain
Ribonucleoprotein complex biogenesis
Mitochondrion organization
Establishment of protein localization to membrane
Protein targeting
Mitochondrial translational initiation
Translational initiation
Nuclear-transcribed mRNA catabolic process
Protein localization to membrane
Mitochondrial translation
Protein targeting to membrane
Translational termination
Purine nucleoside triphosphate metabolic process
Cellular respiration
Nucleoside triphosphate metabolic process
ncRNA processing
Mitochondrial electron transport, NADH to ubiquinone
Ribosome biogenesis
Ribonucleoside triphosphate biosynthetic process
Mitochondrial gene expression
Purine ribonucleoside triphosphate metabolic process
ATP metabolic process
Nucleoside monophosphate metabolic process
Translational elongation
Purine ribonucleoside monophosphate metabolic process
rRNA processing
Ribonucleoside monophosphate metabolic process
Purine nucleoside monophosphate metabolic process
rRNA metabolic process
mRNA metabolic process
Energy derivation by oxidation of organic compounds
RNA catabolic process
Cellular protein complex disassembly
Purine nucleotide metabolic process
Generation of precursor metabolites and energy
mRNA processing
RNA splicing
Ribose phosphate metabolic process
Ribonucleoprotein complex biogenesis
Purine ribonucleoside triphosphate metabolic process
Protein-containing complex assembly
Ribonucleotide metabolic process
Protein complex subunit organization
RNA splicing, via transesterification reactions
Protein-containing complex disassembly
RNA splicing, via transesterification reactions with bulged adenosine as nucleophile
mRNA splicing, via spliceosome
tRNA processing
Mitochondrial transport
Cytochrome complex assembly
Proteasomal protein catabolic process
Ribosome assembly
Cytoplasmic translation
Spliceosomal snRNP assembly
Ribosomal large subunit assembly
Mitotic sister chromatid biogenesis
Proteasome-mediated ubiquitin-dependent protein catabolic process
Mitotic nuclear division
Sister chromatid segregation
Ribosomal small subunit assembly
Ribosomal large subunit assembly
Nuclear division
Regulation of chromosome segregation
Mitotic cell cycle phase transition
Anaphase-promoting complex-dependent catabolic process
Organelle fission
Nuclear chromosome segregation
Protein folding
Response to endoplasmic reticulum stress
Negative regulation of cell cycle process
Regulation of mitotic sister chromatid separation
Mitotic sister chromatid segregation
Regulation of mitotic cell cycle checkpoint
Regulation of chromosome organization
Neutrophil mediated immunity
Establishment of protein localization to mitochondrion
Mitotic cell cycle separation
Release of cytochrome c from mitochondria
Covalent chromatin modification
Histone modification
Protein polyubiquitination
Regulation of GTPase activity


m1
4.11%
95.89%
m2
26.97%
73.03%
m3
42.55%
57.45%
m4
26.74%
73.26%
ccA
28.57%
71.43%
ccB
20%
80%
High-risk group
Low-risk group